Künstliche Intelligenz und Moral

Aufgaben

1 Fassen Sie den vorliegenden Text in eigenen Worten zusammen. (Material)

(30 BE)

2 Vergleichen Sie den Begriff der „moralischen Handlungsfähigkeit“ von Maschinen (Material) mit der moralphilosophischen Konzeption Kants.

(40 BE)

3 „Durch unsere unbeschränkte prometheische[1] Freiheit, immer Neues zu zeitigen[2] […], haben wir uns als zeitliche Wesen derart in Unordnung gebracht, dass wir nun als Nachzügler dessen, was wir selbst projektiert und produziert hatten, mit dem schlechten Gewissen der Antiquiertheit[3] unseren Weg langsam fortsetzen oder gar wie verstörte Saurier zwischen unseren Geräten einfach herumlungern.“

Nehmen Sie unter Berücksichtigung des oben genannten Zitats von Günther Anders aus dem Jahr 1956 Stellung zum Verhältnis von Mensch und Maschine. Gehen Sie dabei auf die Ausführungen Misselhorns und Ihnen bekannte Menschenbilder ein.

(30 BE)

[1] prometheisch – an Kraft und Größe alles überragend; geht auf Prometheus aus der griechischen Mythologie zurück, der den Göttern das Feuer stahl, um es den Menschen zu bringen.
[2] zeitigen – hervorbringen
[3] Antiquiertheit – das nicht mehr Zeitgemäße, Überholte

Material

Catrin Misselhorn: Maschinenethik und „Artificial Morality" – Können und sollen Maschinen moralisch handeln? (2018)

Maschinenethik ist ein neues Forschungsgebiet an der Schnittstelle von Informatik und Philosophie, das die Entwicklung moralischer Maschinen zum Ziel hat. Es geht darum, auf der Grundlage von Computertechnologie Maschinen zu gestalten, die selbst moralische Entscheidungen treffen und umsetzen können. Beflügelt wird dieses Vorhaben von den jüngsten Entwicklungen der Künstlichen Intelligenz. Sollen im Rahmen der Maschinenethik Verfahren der Künstlichen Intelligenz eingesetzt werden, so spricht man analog zu „Artificial Intelligence" (AI) von „Artificial Morality" (AM).

Während AI zum Ziel hat, die kognitiven Fähigkeiten von Menschen zu modellieren oder zu simulieren, geht es bei der AM darum, künstliche Systeme mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten. Dieses Vorhaben wird von einigen euphorisch begrüßt, während andere dadurch einen menschlichen Kernbereich bedroht sehen. […]

Je komplexer und autonomer künstliche Systeme werden, desto eher müssen sie in der Lage sein, ihr Verhalten in einem gewissen Rahmen selbst zu regulieren. Das bringt es mit sich, dass sie auch in Situationen geraten, die moralische Entscheidungen verlangen. […]

Doch selbst wenn man zugesteht, dass es in vielen Anwendungsbereichen sinnvoll wäre, wenn Maschinen moralisch handeln könnten, ist damit noch nicht geklärt, ob sie dazu auch in der Lage sind. Die erste Frage ist, ob autonome Systeme überhaupt handeln können. Die zweite ist, ob die Handlungen künstlicher Akteure als moralisch gelten können.

Die Problematik der grundsätzlichen Handlungsfähigkeit lässt sich innerhalb der philosophischen Handlungstheorie entlang zweier Achsen beschreiben: der Fähigkeit, als selbstursprüngliche Quelle des eigenen Tuns zu fungieren, sowie der Fähigkeit, sich an Gründen zu orientieren. Beide Fähigkeiten müssen als graduelle Attribute begriffen werden, das heißt, sie kommen verschiedenen Arten von Akteuren in unterschiedlichem Maße zu. Der Begriff der Selbstursprünglichkeit wurde von der philosophischen Tradition teilweise im Sinn der Akteurskausalität verstanden, das heißt, dass eine Handlung von einem Akteur ohne vorhergehende Ursache initiiert wird. Ein metaphysisch so anspruchsvoller und umstrittener Begriff der Selbstursprünglichkeit ist jedoch nicht zwingend. Man kann eine einfache Form der Selbstursprünglichkeit auch dann als gegeben sehen, wenn ein System mit der Umwelt interagiert (*Interaktivität*), dabei eine gewisse Anpassungsfähigkeit an sich ändernde Bedingungen aufweist (*Adaptivität*) und in der Lage ist, eine Aktivität ohne direkte menschliche Intervention aufzunehmen (*basale Autonomie*).

Über eine solche primitive Form der Selbstursprünglichkeit können auch Maschinen verfügen, insbesondere solche, die von Computern gesteuert werden. Zwar gibt ein Programm vor, wie sich eine Maschine zu verhalten hat, aber im Einzelfall agiert sie, ohne dass ein Mensch eigens eingreift. Werden Verfahren der KI, beispielsweise Maschinelles Lernen, eingesetzt, so ist es sogar die Aufgabe der Maschine, das moralisch angemessene Verhalten selbst aus den Daten zu erschließen.

Die zweite Achse, die Fähigkeit, sich an Gründen zu orientieren, hängt eng mit der Möglichkeit zusammen, Informationen zu verarbeiten. Dem klassischen Humeschen[1] Modell der Handlungsfähigkeit zufolge besteht der Grund einer Handlung in der Kopplung einer Überzeugung mit einer Pro-Einstellung, beispielsweise einem Wunsch: Ich gehe in die Bibliothek, weil ich ein bestimmtes Buch ausleihen will und der Überzeugung bin, dass es in der Bibliothek vorhanden ist.

---

[1] David Hume (1711–1776) – schottischer Philosoph; bedeutender Vertreter der Aufklärung; wird der philosophischen Strömung des Empirismus zugerechnet

Hinzu kommt nach manchen Ansätzen eine Intention, die dafür verantwortlich ist, dass der Wunsch auch mithilfe eines Plans in die Tat umgesetzt wird. […]

Ein künstliches System kann als funktional äquivalent zu einem menschlichen Akteur verstanden werden, wenn es über Zustände verfügt, denen eine analoge Funktion zukommt, wie Meinungen, Wünschen und Intentionen beim Menschen. […] Das ist ausreichend, um ihnen in einem funktionalen Sinn die Fähigkeit zuzuschreiben, aus Gründen zu handeln. Künstliche Systeme, die zu selbstursprünglichem Handeln aus Gründen in der Lage sind, können als Akteure gelten.

Moralische Handlungsfähigkeit wiederum liegt in einfacher Form vor, wenn die Gründe, nach denen ein System handelt, moralischer Natur sind. Dies ist auf einer rudimentären Ebene schon dann gegeben, wenn ein System über Repräsentationen moralischer Werte[2] verfügt, die die zuvor angegebenen basalen Bedingungen für das Handeln aus Gründen erfüllen (das heißt, es gibt funktionale Äquivalente zu moralischen Überzeugungen, moralischen Pro-Einstellungen und Intentionen). Wenn ein System beispielsweise den Wert der Patientenautonomie[3] als Pro-Einstellung besitzt und zu der Überzeugung kommt, dass dieser Wert in einer bestimmten Situation verletzt wird, dann wird es versuchen, so auf die Situation einzuwirken, dass dieser Wert wieder realisiert wird. Vollumfängliche moralische Handlungsfähigkeit, wie sie Menschen typischerweise besitzen, kommt künstlichen Systemen allerdings nicht zu. Zum einen ist der Einsatzbereich von Maschinen mit Moral normalerweise auf einen bestimmten Anwendungskontext beschränkt, die menschliche Moralfähigkeit umfasst jedoch potenziell jeden beliebigen Bereich des Lebens.

Zudem verfügen künstliche Systeme bislang nicht wirklich über Bewusstsein und Willensfreiheit.

Catrin Misselhorn: Maschinenethik und „Artificial Morality“ – Können und sollen Maschinen moralisch handeln?, in: Aus Politik und Zeitgeschichte, 6–8/2018, URL: https://www.bpb.de/apuz/263684/koennen-und-sollen-maschinen-moralisch-handeln (abgerufen am 02.01.2021).

Hinweise
Catrin Misselhorn (*1970) ist Professorin für Philosophie an der Georg-August-Universität in Göttingen. Ihre Schwerpunkte sind Wissenschaftstheorie und Technikphilosophie.

Günther Anders war ein deutsch-österreichischer Philosoph (1902–1992), der sich mit den ethischen und technischen Herausforderungen seiner Zeit beschäftigte. Sein Hauptthema war die Zerstörung der Humanität.

[2] Repräsentationen moralischer Werte – *hier*: ein moralischer Programmcode im Computer
[3] Patientenautonomie bedeutet, dass Patientinnen und Patienten das Recht haben, über alle bei ihnen vorgenommenen Therapien und Behandlungen selbst zu entscheiden.